## Transform and Enrich Data in Oracle Analytics

Before You Begin

This lab shows you how to perform a set of manual and recommended data preparation actions to your dataset.

Background

Preparing and cleansing your data is an important step before visualizing a dataset. For example, the set might have sensitive data such as customers' social security numbers that you don't want to expose. You can hide or transform all characters of the social security number column, remove columns from a dataset, or extract portions of a data column to create a new column that contains the extracted data.

You can use the recommendations and available data preparation options in Oracle Analytics to improve data quality.

In this tutorial, you use a spreadsheet as the data source. You can add spreadsheet files that have the XLSX extension and that are no larger than 100 MB. You can also use comma-separated value (CSV) and text (TXT) files to create datasets. You can perform data preparation actions on supported data sources.

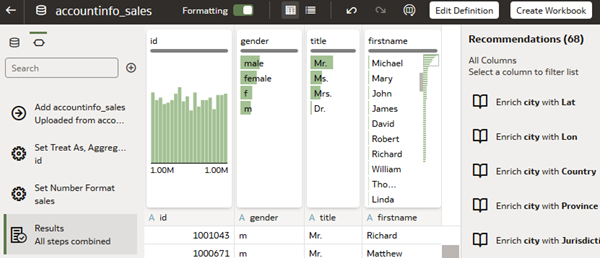
What Do You Need?

* Access to Oracle Analytics Cloud or Oracle Analytics Desktop
* Download the accountinfo\_sales.xlsx file to your computer.

Create a Data Source

Oracle Analytics displays recommendations for the data, by column, in the dataset. In this tutorial, you accept some of the recommendations that are relevant for your analysis. You can also implement transformation changes for data in columns that don't have specific recommendations.

1. Sign in to Oracle Analytics.
2. On the **Home** page, click **Create**, and then click **Dataset**.
3. In **Create Dataset**, click **Drop data file here or click to browse**, select the accountinfo\_sales.xlsx file, and then click **Open**.
4. In Create Dataset Table from accountinfo\_sales.xlsx, click **OK**.
5. In the Join Diagram, click the **accountinfo\_sales** tab.
6. In the Transform Editor, select the **id** column. In Properties, click **Measure** in the Treat As row, and then select **Attribute**.
7. Select the **Sales** column. In Properties, click **Number Format** Number edit icon. In the Number Format row, click **Number**, and then select **Currency**.
8. Click **Save**. In Save Dataset as, enter accounting\_salesinfo, and then click **OK**.

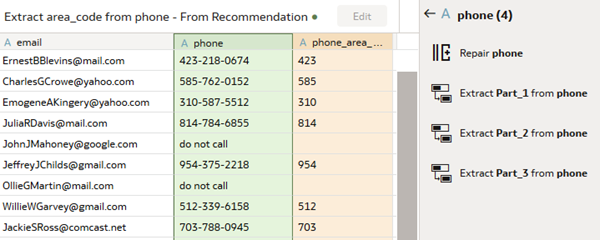


Extract Data from a Column

When you extract data from a column, a new column is created that contains the extracted data. In this section, you extract the area code from phone numbers that use the North American Numbering Plan.

1. In the accountinfo\_sales dataset, click **Toggle Quality Tiles** Toggle Quality Insights icon to close the insights over each column.
2. Select the **phone** column. In the **Recommendation** list, click **Extract area code from phone**.

Oracle Analytics adds an area code column to the dataset.



Conceal Sensitive Customer Data

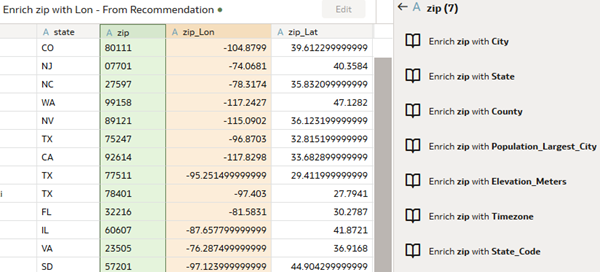
To comply with security policies for sensitive data, you can obfuscate all or a portion of the data in a column. If some users need to see the sensitive data, you can create a duplicate dataset containing the sensitive data.

1. Select the **ccnumber** column.
2. In the **Recommendations** list, click **Obfuscate First 12 Digits of ccnumber**.
3. Select the **ssn** column. In the **Recommendations** list, click **Obfuscate First 5 digits of ssn**.



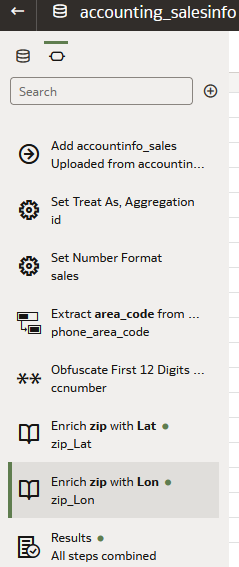
Enrich Data with Geographic Coordinates

1. Select the **zip** column. In **Recommendations**, click **Enrich zip with Lat** (latitude).
2. In **Recommendations**, click **Enrich zip with Lon** (longitude).



1. Click **Save** Save icon.

Your changes are listed in the Preparation Script Preparation Script icon pane, and then applied when you save the dataset.



Inspect the Dataset

In this section, you review the changes implemented in the dataset.

1. Click **Go back** Go back icon.
2. On the Home page, select the **accountinfo\_sales** dataset, click the **Actions menu** Actions menu icon, and then select **Inspect**.
3. In the dataset page, click **Data Elements** to view the columns added to the dataset.

